Logistic Regression

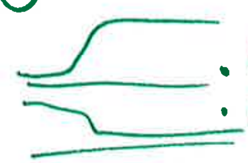
Response has 2 categories, by convention numbered 0 and 1

Ex: $y_i = \begin{cases} 1 & \text{if a crab number } i \text{ is orange species of genus Leptograpsus} \\ 0 & \text{if crab \# } i \text{ is blue species of genus Leptograpsus} \end{cases}$

x_i = frontal lobe size of crab i in mm.

Model says that the prob. of being in class 1 is

$$f_1(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$



- Increasing if $\beta_1 > 0$
- Decreasing if $\beta_1 < 0$

Note this is between 0 and 1 as required

Since $f_0(x) + f_1(x) = 1$,

$$f_0(x) = 1 - f_1(x) = 1 - \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{\beta_0+\beta_1 x}}$$

Decision boundary: $f_1(x) = f_0(x) = 0.5$

$$\Rightarrow 0.5 = \frac{1}{1+e^{\beta_0+\beta_1 x}} \quad \Rightarrow 0.5e^{\beta_0+\beta_1 x} + 0.5 = 1$$

$$\Rightarrow e^{\beta_0+\beta_1 x} = 1$$

$$\Rightarrow \beta_0 + \beta_1 x = 0$$

$$\Rightarrow x = \frac{-\beta_0}{\beta_1}$$

Interpretation of β_1

Def: Odds that $Y=1$ =

$$\text{Odds}(Y_i=1|X_i) = \frac{P(Y_i=1|X_i)}{P(Y_i=0|X_i)}$$

Examples:

- If $P(Y_i=1|X_i)=0.75$, $\text{Odds}(Y_i=1|X_i) = \frac{0.75}{0.25} = 3$
- If $P(Y_i=1|X_i) = 0.5$, $\text{Odds}(Y_i=1|X_i) = \frac{0.5}{0.5} = 1$
- If $P(Y_i=1|X_i) = 0.1$, $\text{Odds}(Y_i=1|X_i) = \frac{0.1}{0.9} = 1/9$

Odds in a logistic regression model:

$$\text{Odds}(Y_i=1|X_i) = \frac{P(Y_i=1|X_i)}{P(Y_i=0|X_i)} = \frac{\left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right)}{\left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}}\right)}$$

$$= e^{\beta_0+\beta_1 x_i}$$

$$\text{log-odds} = \log\{\text{Odds}(Y_i=1|X_i)\} = \log\{e^{\beta_0+\beta_1 x_i}\} = \beta_0+\beta_1 x_i$$

Interpretation in terms of odds:

If $x_i^{(new)} = x_i+1$, ~~log~~ odds is

$$e^{\beta_0+\beta_1 x_i^{(new)}} = e^{\beta_0+\beta_1(x_i+1)} = e^{\beta_0+\beta_1 x_i}\cdot e^{\beta_1}$$

Increasing x_i by 1 unit leads to a multiplicative change in the odds of e^{β_1}

Interpretation in terms of log odds:

Increasing x_i by 1 unit leads to an additive change in the log odds of β_1.